Characterization of a Novel Integrative Element, ICESt1, in the Lactic Acid Bacterium Streptococcus thermophilus
Vincent Burrus, Yvonne Roussel, Bernard Decaris, Gérard Guédon

To cite this version:
Vincent Burrus, Yvonne Roussel, Bernard Decaris, Gérard Guédon. Characterization of a Novel Integrative Element, ICESt1, in the Lactic Acid Bacterium Streptococcus thermophilus. Applied and Environmental Microbiology, American Society for Microbiology, 2000, 66 (4), pp.1749 - 1753. <hal-01659518>

HAL Id: hal-01659518
https://hal.archives-ouvertes.fr/hal-01659518
Submitted on 8 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Characterization of a Novel Integrative Element, ICESt1, in the Lactic Acid Bacterium Streptococcus thermophilus

VINCENT BURRUS, YVONNE ROUSSEL,† BERNARD DECARIS,* AND GÉRARD GUEDON

Laboratoire de Génétique et Microbiologie, INRA UA952, Faculté des Sciences, Université Henri Poincaré (Nancy 1), 54506 Vandoeuvre-lès-Nancy, France

Received 17 September 1999/Accepted 6 January 2000

The 35.5-kb ICESt1 element of Streptococcus thermophilus CNRZ368 is bordered by a 27-bp repeat and integrated into the 3’ end of a gene encoding a putative fructose-1,6-biphosphat aldolase. This element encodes site-specific integrase and excisionase enzymes related to those of conjugal transposons Tn5276 and Tn5252. The integrase was found to be involved in a site-specific excision of a circular form. ICESt1 also encodes putative conjugative transfer proteins related to those of the conjugal transposon Tn916. Therefore, ICESt1 could be or could be derived from an integrative conjugative element.

Cocultures of various lactic acid bacteria are used during the manufacture of dairy products. Sequence comparisons and hybridizations reveal that horizontal transfers between a large array of species of lactic acid bacteria have occurred, most likely during dairy cocultures (13, 32). The most convincing evidence indicates that insertion sequences IS1191, IS981, IS1, and IS1194 (4, 5, 14) and some open reading frames (ORFs) involved in exopolysaccharide synthesis (6) or in restriction-modification (24) were transferred between the lactic acid bacteria Streptococcus thermophilus and Lactococcus lactis in cocultures used during cheese manufacture. However, the mechanism of genetic exchange between these two species remains unknown, and no conjugative element has been previously characterized in S. thermophilus.

Cloning of varIC and localization of its limits. The Sm4 fragment of the S. thermophilus CNRZ368 chromosome was previously found to contain the 35-kb variable region varIC, which was absent from the corresponding chromosomal fragments of strains A054 and NST2280 (28). A region containing an IS1191 copy inserted in a truncated IS981 element (14) was cloned and found to be included in varIC (28). Chromosome walking using a AGEM11 genomic library of CNRZ368 (25) was performed to isolate recombinant λ bacteriophages overlapping the varIC region. Their inserts were subcloned in pBC KS+ and used as hybridization probes on A054 and NST2280 DNAs. S35, ES27, I132.3, ES13, and SC02 fragments hybridized to A054 and NST2280 DNAs. On the contrary, all of the probes covering the 35.5-kb region (except IS1191 and IS981) and located between the HindIII sites H_{L} and H_{R} (Fig. 1) did not hybridize to A054 and NST2280 DNAs (data not shown). Furthermore, CNRZ368, A054, and NST2280 showed identical restriction maps in regions located to the left of the HindIII site H_{L} and to the right of the HindIII site H_{R} (Fig. 1). These data indicated that varIC limits are located near these HindIII sites. When ES27 including the left end and ES13 including the right end were hybridized to DNAs of the three strains digested by ClaI, EcoO1099, EcoRI, PstI, or XbaI, they revealed the same fragment from A054 and NST2280, but two different fragments from CNRZ368. Thus, the flanking regions of varIC in CNRZ368 are adjacent to each other in strains A054 and NST2280 (Fig. 1).

Because A054 and CNRZ368 are very closely related to each other, but distantly related to NST2280 (28), the absence of varIC in A054 and NST2280 probably results from an insertion in CNRZ368 rather than from two independent identical deletions in the two other strains.

varIC is bordered by a direct repeat and encodes an integrative system. Sequencing of the varIC limits revealed that the element is bordered by a 27-bp direct repeated sequence (R1) containing a HindIII site (Fig. 2). A 362-bp fragment was obtained by PCRs performed with the DNA of S. thermophilus A054 by using the convergent primers O132.3 (GGACTACT CGGCGGCGG) and O131.2 (TGGTGGATATCAAGAA GC) (Fig. 3). The sequence of this fragment revealed a unique R1 copy identical to those found on either side of varIC in CNRZ368 (Fig. 2). Sequence comparison indicates that R1 direct repeats of CNRZ368 correspond to the boundaries of varIC (Fig. 2).

Two ORFs, int and xis, are located within varIC near the right copy of R1 (Fig. 1 and Table 1). The putative protein encoded by int shows significant similarities to site-specific recombinases belonging to the 6LC3 subgroup of the integrase family (http://members.home.com/domespo/trhome.html). This subgroup includes a large array of integrases of temperate bacteriophages and conjugative transposons of lactic acid bacteria and other gram-positive low-G+C bacteria. The C terminus of Int contains the five amino acids which are perfectly conserved in this family (data not shown) (1, 3, 11). Furthermore, xis, located to the left of the int gene, encodes a small basic protein (MW 9,88) which shows significant similarities to excisionases of two conjugal transposons, Tn5252 of Streptococcus pneumoniae and Tn5276 of Lactococcus lactis (Table 1). int and xis are located at comparable positions in many prophages and conjugal transposons.

Therefore, these ORFs probably encode an integrative system which would mediate excision of varIC by site-specific recombination between the two R1 copies corresponding to the cores of the left and right attachment sites attL and attR. The unique R1 sequence found in A054 would be the attB attachment site used for varIC integration. fda, which flanks the right of varIC (Fig. 1), encodes a putative fructose-1,6-biphosphat aldolase (Table 1). The 3’ end of fda includes 20...
bp of the R1 core of attR (Fig. 2). Thus, var1C integration does not change the sequence of fda. Numerous integrative elements (e.g., prophages or integrative conjugative elements) integrate into the 3’ end of genes encoding tRNAs, their sequences remaining unmodified by the integration (8, 15, 17, 23, 30, 31). Other integrative elements (e.g., most of the conjugative transposons) integrate into several or numerous sites (19, 26). Only a few elements site specifically integrate into the 3’ end of protein-encoding genes. The substitution sequence is then generally similar to the original one (10, 18).

An imperfect 14-bp inverted repeat, R2, is located 29 bp to the right of the 3’ end of the int gene and 21 bp to the left of the R1 core of attR (Fig. 1). The potential stem-loop structure ($\Delta G = -18.8 \text{ kcal} \cdot \text{mol}^{-1}$) (33), preceded by a stretch of A’s and followed by a stretch of T’s, could be used as a p-independent transcription terminator for both int and fda. A perfect 13-bp inverted repeat, R5 ($\Delta G = -14.8 \text{ kcal} \cdot \text{mol}^{-1}$), preceded by a stretch of A’s, is located 2 bp to the left of the core of attL (Fig. 1) and could be used as a transcription termination signal for fda prior to the var1C integration. Therefore, these data suggest that the expression of fda would not be changed after var1C integration.

R3, a perfect 9-bp direct repeat, was found 2 bp downstream from the stop codon of int (Fig. 1). A copy of this 9-bp sequence was also found 148 bp to the right of the R1 core of attL. R6, an imperfect 12-bp inverted repeat, and R4, an imperfect 9-bp inverted repeat, are located 123 and 229 bp to the right of the core of attL, respectively. R2, R3, R4, and R6 could be binding sites for integrase or host-encoded proteins involved in the recombination.

Detection of site-specific recombination products. A nested PCR was performed to amplify the putative junction between the var1C termini, which could result from a site-specific recombination event between the R1 cores of attL and attR. Nested-PCR amplification was performed with the O132.5 (GATGAAATTCACTACATCATC)-O131.5 (CAGGACTCGAT ATTGACA) outer primer pair and the O132.4 (AGTTGAA ACTAGACTCAG)-O131.1 (TTCCAGACATACGGACAT) inner primer pair (Fig. 3A) according to the method described by Manganelli et al. (21). As expected, no product was identified in strain A054 (Fig. 3B), which does not contain var1C.

The sequence of the 536-bp PCR product obtained in CNRZ368 (attl, Fig. 2) is identical to the expected sequence resulting from site-specific recombination between the R1 cores of attL and attR. The PCR product was digoxigenin labelled and hybridized to EcoRI-digested A054 and CNRZ368 chromosomal DNA. As expected, this probe hybridizes with the two fragments containing the var1C termini in CNRZ368, but not with A054 DNA (data not shown). Site-specific excision of var1C in CNRZ368 should also lead to a junction between sequences flanking var1C, identical to that observed in A054. PCR amplification using the O132.3-O131.2 primer pair (Fig. 3A) was performed to detect this junction. PCR products obtained for A054 and CNRZ368 show the same size (Fig. 3B) and restriction map (data not shown).

Detection of these two junction fragments implies in

---

**FIG. 2.** Comparison of the nucleotide sequences of the four attachment sites. attL and attR include, respectively, the left and right terminal of var1C of strain CNRZ368. attB corresponds to the partial sequence of a PCR product obtained from strain A054 with the primer pairs O132.3 and O131.2. attl corresponds to the partial sequence of a nested-PCR product obtained from strain CNRZ368 with the primer pairs O132.5-O131.5 and O132.4-O131.1 (Fig. 3). R1 sequences are written in capital letters. The italic letters correspond to the internal sequence of var1C. Underlined letters indicate the bases that are complementary to the 3’ end of the fda gene encoding fructose-1,6-bisphosphate aldolase. Sequences underlined twice correspond to the HindIII restriction sites included in R1.
The thermosensitive plasmid pNST152 was constructed by subcloning the 754-bp HindIII fragment of pNST131.1 containing a fragment of int (region encoding restriction enzyme). This restriction enzyme from S. thermophilus was designated as Int. Integration probably takes place by crossover events similar to those resulting from the integration of a unique copy of pNST152 within the int gene of CNRZ368. Junction fragments containing attB or attI were not detected in NST1008 by PCR experiments (Fig. 3B), whereas a fragment bearing attR was amplified from NST1008 by using the O131.1 and O131.2 primers (Fig. 3). Therefore, int gene disruption causes the disappearance of the two junction fragments and, therefore, of the covalent circular molecule, showing that this gene is actually involved in var1C excision.

var1C encodes proteins related to those of some conjugative systems. The 5,881-bp region located to the left of the xis ORF start codon was sequenced. Four ORFs have been identified by GeneMark (http://genemark.biology.gatech.edu/GeneMark/) and/or by comparison of the putative translation products with proteins from the EMBL/GenBank databases by using BLASTX and BLASTP (2) (Fig. 1 and Table 1). All of these ORFs are preceded by a suitably located ribosome binding site (RBS) (27), have the same orientation as xis and int, and are spaced by very short sequences (Table 1). Therefore, orfDCBA, xis, and int could be translated from a unique transcript.

orfA and orfD products share significant sequence similarities with proteins involved in conjugative transfer of plasmids from Staphylococcus aureus and Tn916 from Enterococcus faecalis (Table 1). orfC encodes a putative protein weakly related to the translational product of orf15 of the conjugative transposon Tn916. Topology predictions using the HMMTOP server (http://www.enzim.hu/hmmtop/) indicate that the proteins encoded by these two ORFs would be transmembrane proteins with similar tridimensional structures, suggesting that they are actually related. Thus, this region of var1C could encode conjugative functions. Various recently identified elements excise by forming a circular intermediate, promote self-transfer by conjugation into the recipient cell, and integrate by recombination between the specific site of the circular molecule and another site (17, 26, 29, 31). Therefore, the entire var1C sequence could be or could be derived from a site-specific conjugative integrative element. This possible conjugative element, which would be the first isolated in S. thermophilus, was named ICESt1, for conjugative integrative element of S. thermophilus no. 1.

The possible conjugative system of ICESt1 is related to that of Tn916, but not to the system encoded by Tn5252. On the contrary, the ICESt1 excisionase is related only to those of Tn5276 and Tn5252. Moreover, the integrases of ICESt1, Tn5276, and Tn5252 belong to the LC3 integrase subfamily, whereas the integrase of Tn916 belongs to another subfamily (http://members.home.net/domespo/trhome.html). Furthermore, differences in G+C content between the xis and int genes (about 34%) and orfABCD (about 42%) of ICESt1 also suggest that the integration-excision system and the possible conjugative system have different origins or have undergone very different evolutions. A similar structure is observed in Tn916 (about 36% G+C for the xis and int genes versus about 40% G+C for the conjugative system). This suggests that ICESt1 and Tn916 possess a modular structure which results from exchanges or acquisitions of sequences from different sources. This modular structure and evolution are similar to...
TABLE 1. Characteristics of the sequenced ORFs and encoded proteins examined in this study

<table>
<thead>
<tr>
<th>ORF</th>
<th>% G+C</th>
<th>Start</th>
<th>Stop</th>
<th>RBS(^a)</th>
<th>Dist(_t^c)</th>
<th>Dist(_S^d)</th>
<th>Length (no. of amino acids)</th>
<th>Related protein(^e)</th>
<th>Origin</th>
<th>% Identity (no. of amino acids)(^f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fla</td>
<td>38.3</td>
<td>ND(^a)</td>
<td>TAA</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>641</td>
<td>Fructose-1,6-biphosphate aldolase Fba (AJ005697)</td>
<td>Chromosome of S. pneumoniae</td>
<td>80 (77)</td>
</tr>
<tr>
<td>int</td>
<td>33.6</td>
<td>ATG</td>
<td>TAA</td>
<td>TAAGGAGG</td>
<td>7</td>
<td>−1</td>
<td>448</td>
<td>Integrase Int (U93688)(^g)</td>
<td>Pathogenicity island SaPl1 of S. aureus</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Integrase Int (M27965)(^g)</td>
<td>Phage dL54a of S. aureus</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Integrase Int (M62697)(^g)</td>
<td>Phage dadh of Lactobacillus casei</td>
<td>32 (227)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Integrase Int (L27649)(^g)</td>
<td>Conjugative transposon Tn5276 of L. lactis</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Integrase Int (L29324)(^g)</td>
<td>Conjugative transposon Tn5252 of S. pneumoniae</td>
<td>26 (266)</td>
</tr>
<tr>
<td>xis</td>
<td>35.7</td>
<td>ATG</td>
<td>TAA</td>
<td>AAAGGAGT</td>
<td>5</td>
<td>+13</td>
<td>82</td>
<td>Excisionase Xis (L29324)</td>
<td>Conjugative transposon Tn5252 of S. pneumoniae</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Excisionase Xis (L27649)</td>
<td>Conjugative transposon Tn5276 of L. lactis</td>
<td>41</td>
</tr>
<tr>
<td>orfA</td>
<td>43.3</td>
<td>ATG</td>
<td>TAG</td>
<td>AAAAGAGA</td>
<td>4</td>
<td>+25</td>
<td>370</td>
<td>Putative transfer protein TraG (AR051917)</td>
<td>Conjugative plasmid pSK41 of S. aureus</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Putative transfer protein TraG (L11998)</td>
<td>Conjugative plasmid pGO1 of S. aureus</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Putative lipoprotein Isp (U31811)</td>
<td>Chromosome of Streptococcus pyogenes D471</td>
<td>57 (325)</td>
</tr>
<tr>
<td>orfB</td>
<td>37.8</td>
<td>ATG</td>
<td>TAA</td>
<td>AGAGGAGA</td>
<td>5</td>
<td>+1</td>
<td>74</td>
<td>No similarity</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>orfC</td>
<td>41.5</td>
<td>ATG</td>
<td>TAG</td>
<td>TTAGGAGG</td>
<td>7</td>
<td>+11</td>
<td>626</td>
<td>Putative membrane protein Orf15 (U09422)</td>
<td>Conjugative transposon Tn916 of E. faecalis</td>
<td>18 (267)</td>
</tr>
<tr>
<td>orfD</td>
<td>42.1</td>
<td>ATG</td>
<td>TAG</td>
<td>AAAAGAGA</td>
<td>4</td>
<td>ND</td>
<td>834</td>
<td>Putative transfer protein Orf16 (U09422)</td>
<td>Conjugative transposon Tn916 of E. faecalis</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unknown protein YddE (AB001488)</td>
<td>Conjugative transposon Tn916 of E. faecalis</td>
<td>30</td>
</tr>
</tbody>
</table>

\(^a\) ORFs are listed from the right to the left of the map.
\(^b\) The RBS consensus sequence of the gram-positive low-G+C bacterium Bacillus subtilis is AAAGGAGG.
\(^c\) Dist\(_t^c\), distance between the RBS and the start codon.
\(^d\) Dist\(_S^d\), distance between the start codon of an ORF and the stop codon of the previous ORF on the map (Fig. 1). A negative value indicates an overlapping of two ORFs.
\(^e\) Functions of proteins and GenBank accession numbers (in parentheses) of nucleotide sequences encoding proteins related to the product of the ORFs sequenced in this study are indicated.
\(^f\) Identities stretch over the entire length of each of the amino acid sequences of proteins encoded by ICESt1, except when indicated in parentheses.
\(^g\) ND, not determined.

Many other related integrases were found in databases, but in this table, we have only indicated a selection of the ones more related to the integrase encoded by var1C.